DOCUMENT RESUME

ED 187 750 TM 800 252

AUTHOR House, Gary D.
TITLE An Application of Regression Theory to the Evaluation of Pupil Identification Procedures for ESEA Title I Programs.
PUB DATE Apr 80
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Compensatory Education; Elementary Education; Elementary School Mathematics; Identification; Reading Achievement; *Regression (Statistics); *Screening Tests; *Test Coaching
IDENTIFIERS *Elementary Secondary Education Act Title I; Iowa Tests of Basic Skills

ABSTRACT

This paper analyzes data from three administrations (spring and fall 1977 and spring 1978) of reading and mathematics achievement tests across grades three through eight. The study asks two major questions: (1) how accurately are students identified for Elementary Secondary Education Act Title I services based upon the first test administration?; and (2) how accurately are students not identified based upon that test? Through proportional analysis, and comparisons of actual test performances to regression-predicted test performances for several subgroups within the tested population, it was found that students scoring low on the selection test were much more likely to score low on subsequent tests than were high scoring pupils likely to continue to score high. Regression-predicted performance was much more accurate for low scoring pupils than for high scoring pupils. Coaching is strongly suggested by these data. Implications include that the school system in question needs to change its project identification procedures and needs to monitor its testing program. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187750

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

An Application of Regression Theory to the Evaluation of Pupil Identification Procedures for E.S.E.A. Title I Programs

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. House

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by
Gary D. House
Parkway School District
455 North Woods Mill Road
Chesterfield, Missouri 63017

Paper presented at the Annual Meeting
of the American Educational Research Association,
Boston, Massachusetts, April, 1980, Session No. 2.13

TM 800252

Introduction

A large midwestern school system spent approximately $10 million serving more than 20 thousand pupils from preschool through high school age in ESEA, Title I reading and/or arithmetic treatments during the 1977-1978 academic year. Virtually all pupils served were identified for treatment through a single spring administration of the Iowa Tests of Basic Skills. On-level administrations of these tests produced normal curve equivalent (n.c.e.) scores in reading comprehension and total arithmetic subtests, which were later used on identifying educationally disadvantaged pupils for remedial reading and remedial arithmetic programs respectively. A pupil enrolled below grade seven during the year of treatment (1977-78) was categorized identified for treatment if his prior spring ITBS n.c.e. score in reading or arithmetic was less than 46. Pupils in grades 7 and 8 had to have scored 42 or lower on the appropriate prior spring test in order to be served during 1977-78.

The 1977-1978 academic year was the first in which the school system was implementing U.S.O.E., Title I Evaluation Model A_1; therefore, in addition to the usual spring, 1978 administration of the ITBS, a fall, 1977 administration of the same form and level of the reading comprehension and arithmetic subtests was required. In no case was the fall administration to be used for the selection of pupils into Title I treatments. In order to protect against misuse of fall results, they were not released until well into the second semester, and only grade equivalent scores were released at that time.

A complete testing sequence for 1977-78 was:

1. Spring, 1977 testing (ITBS, Form 5) for general purposes and for selection into 1977-78 Title I treatments.
2. Fall, 1977 testing (ITBS, Form 6) for Model A, pretest purposes only.
3. Spring, 1978 testing (ITBS, Form 6) for general purposes, for posttesting in the Model A_1 evaluation design, and for selection purposes related to 1978-79 Title I treatments.

Not all pupils were eligible for Title I services, even though all were tested according to the same testing sequence. In order to be Eligible for Title I treatment, it was necessary for pupils to reside within the boundaries of neighborhood schools meeting state guidelines for economic deprivation.

The entire population of students could be subdivided into four distinct categories:

1. Pupils who were Ineligible and Not Identified for Title I treatment. These pupils lived in neighborhoods not categorized as economically deprived and they had test scores higher than those which would identify them for Title I services.
2. Pupils who were Eligible and Not Identified. Pupils who resided in Title I school attendance areas but who had relatively high prior spring test scores fell into this category.
3. Pupils who were Ineligible and Identified. Pupils in this category were from non Title I attendance areas but had qualifying test scores to be offered treatment if such treatment had been available.
4. Eligible and Identified pupils. These pupils lived in Title I neighborhoods and scored low enough to be considered for treatment.

Statement of the Problem

The testing program of the school system in which the current study was conducted had been loosely run for many years. No test security procedures were in effect, and according to records of test purchase orders and current warehouse inventories, thousands of test booklets had been lost. Rumor had it that some pupils, especially in one subdistrict, had been regularly coached during the normal spring census testing process of the district. The subdistrict in question was one in which every school was eligible for Title I services.

The school system had, since its initial use of census norm-referenced testing, placed a good deal of emphasis on test results, and had regularly encouraged teachers to work hard at raising test scores.

Each year, rosters of pupil grade-equivalent scores and profiles of classroom, school, and district grade-equivalent performance had been published. The five subdistricts had often been compared to one another on test results, and an air of competition had been apparent among the professional community.

This study was undertaken in order to determine whether or not there was evidence among available test data used in implementing one school year's programs that coaching had taken place. Especially important to the study were the ramifications of the hypothesized condition for the operation of Title I programs. That is, if coaching had taken place, then many pupils who might have otherwise qualified for Title I services may have been denied those services inappropriately.

According to Campbell and Stanley (1963), Campbell and Erlebacher (1970), Campbell (1969), Hopkins (1977), and others, the means of samples of subjects, drawn on the basis of being different from the population mean, will always regress toward the population mean on subsequent administrations of imperfectly correlated tests.

The distance between the population mean and a sample mean on a subsequent test administration can be predicted, given no intervening special treatment or practice for the sample (Hopkins, 1977). The prediction is based upon the correlation between the tests and can be expressed as:

$$\hat{Y}_s = \bar{Y}_p - r(\bar{Y}_p - \bar{X}_s) + E$$

where, $\hat{Y}_s$ = predicted posttest mean for the sample

$\bar{Y}_p$ = observed posttest mean for the population

$\bar{X}_s$ = observed pretest mean for the sample

r = pretest to posttest correlation coefficient for the population

E = error

Although not all authors agree as to how much regression effect would continue through two, three, four or more subsequent test administrations (see Tallmadge, 1976) there is general agreement that groups selected on the basis of their differences from their population mean will regress toward the population mean on at least the first subsequent administration of an imperfectly correlated test. This study makes use of the above stated regression theory to assess the extent to which some real data conform to that theory and are, therefore, reasonable indices of the reading and arithmetic achievement of sampled subjects.

Two major questions are addressed:

1. How accurate was the single selection test in identifying categories of pupils for Title I treatment?
2. How accurate was the single selection test in not identifying pupils for Title I treatment?

It was, from the outset, recognized that this study would be entirely ex post facto, and that any evidence presented in answer to its questions would be entirely circumstantial. Yet, the implications for finding results which might point to poor pupil identification procedures could be useful to the school system in which the study took place as decisions with regard to this important function are made. Other school systems, faced with similar problems, may also be interested in the study.

The Hypothesis

As this study was designed, it was reasoned that if systematic coaching, or some other error factor, was operating during the regular spring testing program in order to artificially raise spring test scores, the same phenomenon would probably not occur in the new fall testing sequence, and thus, fall scores would, on average, be somewhat lower systemwide than either the previous spring scores or the following spring scores.

Since rumor had it that coaching occurred more often in the generally lower scoring Title I eligible districts, it was further reasoned that the previous spring test scores for relatively high scoring Title I eligible pupils would drop more in the fall than would the scores of all other groups, and more than the population as a whole. The most critical test for this expectation is that the fall scores of relatively high scoring Title I eligible pupils would drop more in the fall than would the scores of relatively high scoring pupils who were ineligible for Title I services.

On the other hand, it was expected that the fall test performance for relatively low scoring Title I eligible pupils would not change substantially in the fall, nor would it change very much in relation to the scores of relatively low scoring pupils who were not Title I eligible.

Likewise, it was speculated that the high scoring Title I eligible pupils would perform on subsequent tests substantially differently from predictions based upon statistical regression; whereas, other groups of pupils were expected to perform within reasonable error margins of predicted mean values. This expectation was based upon an assumption that regression theory is generally valid, and that systematic or regular observations of departure from

expectations based upon this theory would be evidence of a rather significant intervention of some sort.

Given the recognition that this study was of necessity entirely ex post facto, with no controls exercised over any variables, no firm conclusions as to the causes of any confirmed hypotheses are possible. The study was undertaken to determine whether or not evidence which would not rule out coaching was present. It was understood that given an inability to rule out coaching, further study, carefully designed prior to future test administrations, would be necessary. Keeping the above in mind, the following hypothesis was generated:

The possibility of a widespread occurrence of spring test coaching, especially in Title I eligible schools, cannot be ruled out if all of the following conditions are met:

1. Fall test performance for the entire tested population is lower at each grade level than is either the prior spring performance, or the subsequent spring performance.

2. Fall test results for eligible and not identified pupils drop more than any other group, and more than the population as a whole.

3. Fall test performance for eligible and identified pupils does not change substantially from previous spring levels, nor does it change substantially from that of their Title I ineligible, relatively low scoring counterparts.

4. Performance on tests subsequent to the previous spring selection test does vary significantly from statistical expectations for Title I eligible and not identified pupils. The mean test performance for all other groups does not differ substantially from statistical predictions.

Methods

A file containing all individual test results from spring, 1977, fall, 1977, and spring, 1978, for pupils in grades three through eight, was created. The study was limited to these grades because a single test battery (ITBS) was used throughout, with the exception of the spring, 1977 scores of third graders during 1977-78. Third grade pupils were in grade two during the previous spring, where the Primary Battery of the ITBS was administered in Title I eligible schools only. At no other grade level was a test administered to all pupils. Therefore, no other grades were suitable for inclusion into this study's data base.

Only those pupils having valid test scores across all three test administrations were included in any of the analyses.

Populations of pupils by grade level were partitioned into four distinct groups based upon their spring, 1977 test scores in reading and/or in arithmetic:

Group 1: Ineligible and Not Identified Pupils

Group 2: Eligible and Not Identified Pupils

Group 3: Ineligible and Identified Pupils

Group 4: Eligible and Identified Pupils

Analysis of the data occurred in two steps. The first step consisted of an examination of proportions of pupils who scored at or below selection criteria in reading and/or arithmetic for Title I program participation (i.e., were identified). Proportions of pupils identified on the selection test (spring, 1977), the subsequent pretest (fall, 1977), and the posttest (spring, 1978) were examined for among-group differences.

Secondly, means of all tests were calculated for each group and the entire population. Population correlation coefficients were determined between selection tests and pretests, and between pretests and posttests, by subtest and grade level. These data were used to calculate predicted means for each group on the pretest and posttest administrations in both subject matter areas. Predicted and observed means for each group and observed means for the population were then plotted and examined against the criteria mentioned under the hypothesis.

Results

Tables one and two contain information on the size of each group by grade level and on the proportion of those groups who met identification criteria at each of the three testing points. Table one contains data on the reading subtests; table two contains data on the total arithmetic subtests.

Table one shows that of the eligible pupils from grades three through eight who were identified in reading based upon the selection test, 91 percent continued to be identified in the fall, and 89 percent continued to be identified the following spring. By grade level, the proportions identified in the fall ranged from 89 percent in grade three to 94 percent in grade eight. The range of identified pupils the following spring ranged from 81 percent in grade seven to 95 percent in grade four.

Of those pupils who were ineligible but had identifying selection test scores, 81 percent across the grades were still identified based on pretest results, and 81 percent were identified the following spring. By grade level, the pretest-identified proportions' range was from 76 percent in grade seven to 86 percent in grade four. Although the proportions of identified pupils who remained identified on the two tests following the selection procedure were slightly different between the two groups of identified pupils, the tendency of the three tests to consistently identify the vast majority of pupils was quite similar.

Eligible pupils who scored higher than the selection criterion on the spring, 1977 test were not consistently found to be not identified on the

Table 1

Percents of Samples of Pupils Who Were Classified As Identified at Three Testing Points.

READING

Grade	Test	Eligible and Identified	Ineligible and Identified	Eligible and Not Identified	Ineligible and Not Identified
		N = 926	N = 95	N = 661	N = 32
	Selection	100%	100%	0%	0%
3	Pretest	89%	78%	62%	72%
	Posttest	90%	86%	51%	59%
		N = 1630	N = 485	N = 908	N = 772
	Selection	100%	100%	0%	0%
4	Pretest	91%	86%	52%	30%
	Posttest	95%	89%	57%	35%
		N = 1854	N = 583	N = 563	N = 670
	Selection	100%	100%	0%	0%
5	Pretest	92%	83%	45%	24%
	Posttest	91%	82%	41%	22%
		N =1950	N = 654	N = 497	N = 705
	Selection	100%	100%	0%	0%
6	Pretest	90%	82%	46%	22%
	Posttest	89%	84%	40%	21%
		N = 1655	N = 690	N = 659	N = 808
	Selection	100%	100%	0%	0%
7	Pretest	90%	76%	46%	22%
	Posttest	81%	76%	29%	15%
		N = 1627	N = 621	N = 873	N = 994
	Selection	100%	100%	0%	0%
8	Pretest	94%	81%	49%	21%
	Posttest	90%	74%	34%	16%
		N = 9642	N = 3128	N = 4161	N = 3981
	Selection	100%	100%	0%	0%
All	Pretest	91%	81%	50%	24%
	Posttest	89%	81%	42%	22%

Table 2

Percents of Samples of Pupils Who Were Classified as Identified at Three Testing Points.

ARITHMETIC

Grade	Test	Eligible and Identified	Ineligible and Identified	Eligible and Not Identified	Ineligible and Not Identified
		N = 807	N = 94	N = 773	N = 28
	Selection	100%	100%	0%	0%
3	Pretest	90%	81%	63%	25%
	Posttest	87%	77%	56%	36%
		N = 1524	N = 447	N = 1070	N = 782
	Selection	100%	100%	0%	0%
4	Pretest	92%	82%	48%	28%
	Posttest	90%	82%	41%	27%
		N = 1385	N = 450	N = 848	N = 767
	Selection	100%	100%	0%	0%
5	Pretest	90%	81%	50%	28%
	Posttest	82%	74%	40%	19%
		N = 1644	N = 548	N = 979	N = 829
	Selection	100%	100%	0%	0%
6	Pretest	93%	86%	53%	26%
	Posttest	84%	79%	33%	22%
		N = 1417	N = 596	N = 950	N = 916
	Selection	100%	100%	0%	0%
7	Pretest	89%	82%	43%	22%
	Posttest	76%	75%	28%	14%
		N = 1521	N = 605	N = 1021	N = 999
	Selection	100%	100%	0%	0%
8	Pretest	87%	74%	38%	18%
	Posttest	78%	71%	27%	14%
		N = 8298	N = 2740	N = 5641	N = 4321
	Selection	100%	100%	0%	0%
All	Pretest	90%	81%	48%	24%
	Posttest	83%	76%	37%	19%

subsequent tests. Overall, 50 percent became identified on the pretest and 42 percent on the posttest. The range of proportions on the pretest was from 45 percent at grade five to 62 percent at grade three; on the posttest, the range was from 29 percent at grade seven to 57 percent at grade four.

The ineligible and not identified pupils had different results. Across all six grades, 24 percent became identified on the pretest, and 22 percent were identified on the posttest. The pretest range was from 21 percent in grade eight to 72 percent[1] at grade three. The posttest range was from 15 percent in grade seven to 59 percent in grade three.

Table two shows results, for the arithmetic total subtest, which are similar to those shown for reading. Again, eligible and identified groups have results like, but not identical with ineligible and identified groups. Both sets of identified pupils show the major portion of their subjects continuing to be identified across pretest and posttest periods, but with eligible samples showing the most consistent results.

Dissimilarly, eligible and not identified pupils become identified at about twice the rate of ineligible and not identified pupils.

---

[1] It should be noted that all pupils ineligible in grade three had to have been eligible the previous year in order to have been tested at all, since the only primary-aged pupils tested were in Title I schools.

Figures one through six show plotted observed selection test, pretest, and posttest means in reading for the four named groups and for the population. Observed means are connected by solid lines and the groups are identified by numbers as defined on pages 2 and 8. The population curve is not numbered. Predicted pretest means and predicted posttest means are plotted for each of the four groups, and those predicted means are connected to prior means, from which they were projected, by broken lines.

Figures seven through twelve contain similar data based upon arithmetic total subtest results.

Figures one through twelve show a consistent decline in pretest performance compared to selection test performance. This decline in population performance is, in every instance, followed by an increased mean performance on the posttest.

With exceptions in grade three, the ineligible and not identified group (group 1) contained the overall highest achieving pupils and this group consistently performed at or better than predictions.

In every instance, group two, the eligible and not identified pupils, achieved pretest means much lower than predicted, and they achieved posttest means much higher than predicted. This trend toward lower, then higher achievement than predicted was not observed for any other group.

Observed means for the two low scoring groups varied around predictions unsystematically, but remained well below population values.

For all groups, prediction accuracy improved as grade levels increased, but predictions were always more accurate for low scoring than for high scoring subjects, and were least accurate for group two subjects.

Figure 1
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 3

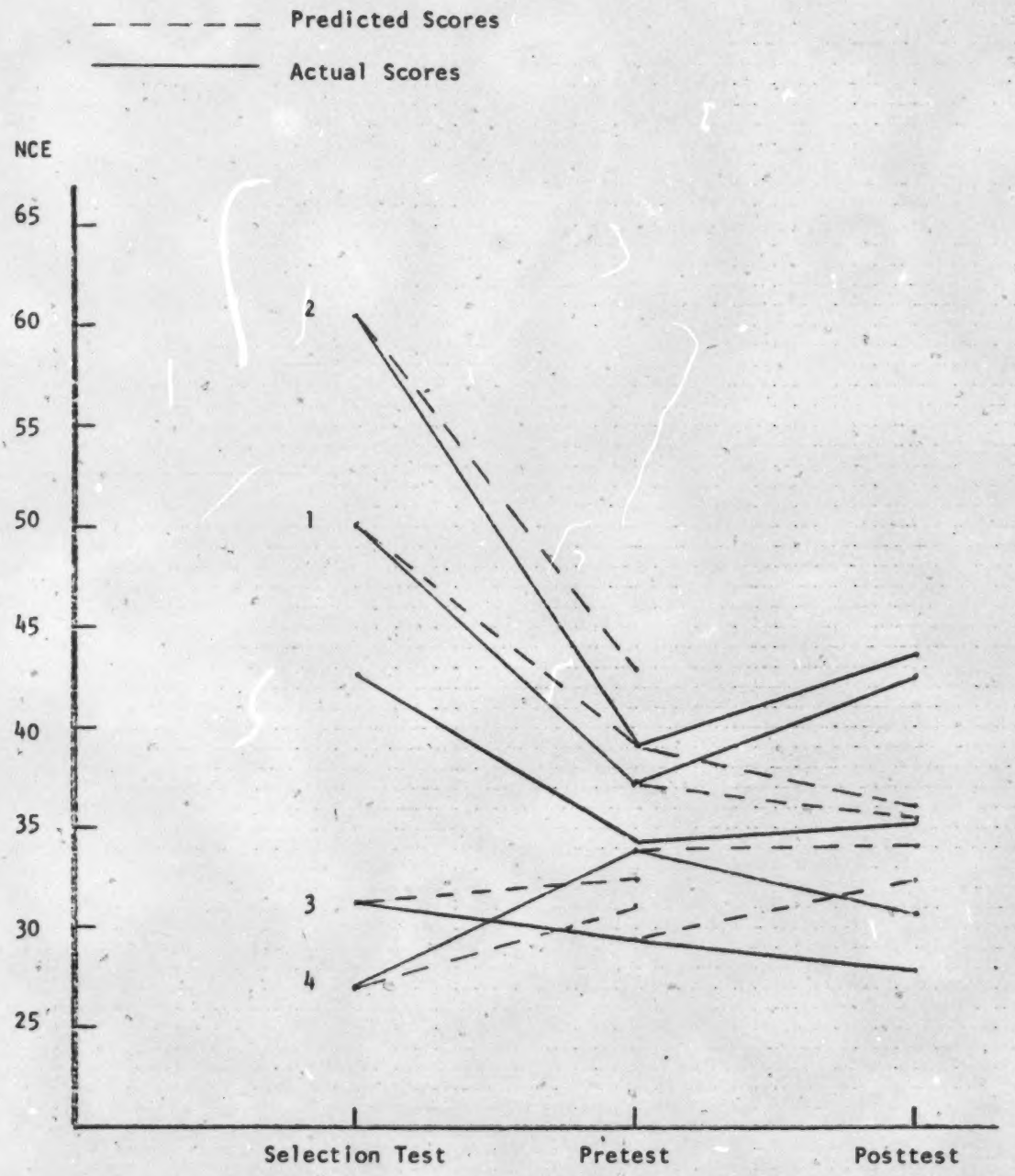


Figure 2
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 4

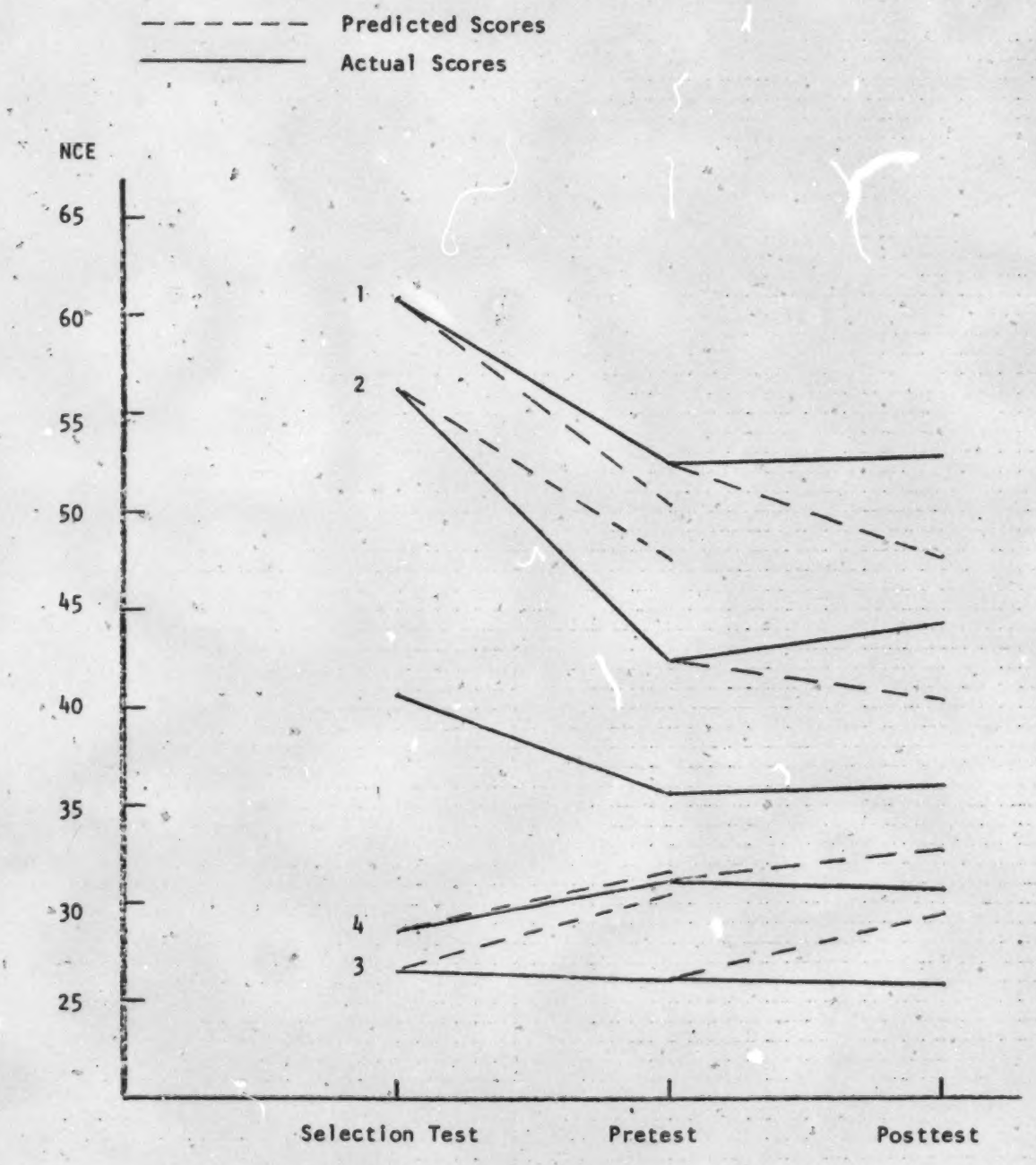


Figure 3
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 5

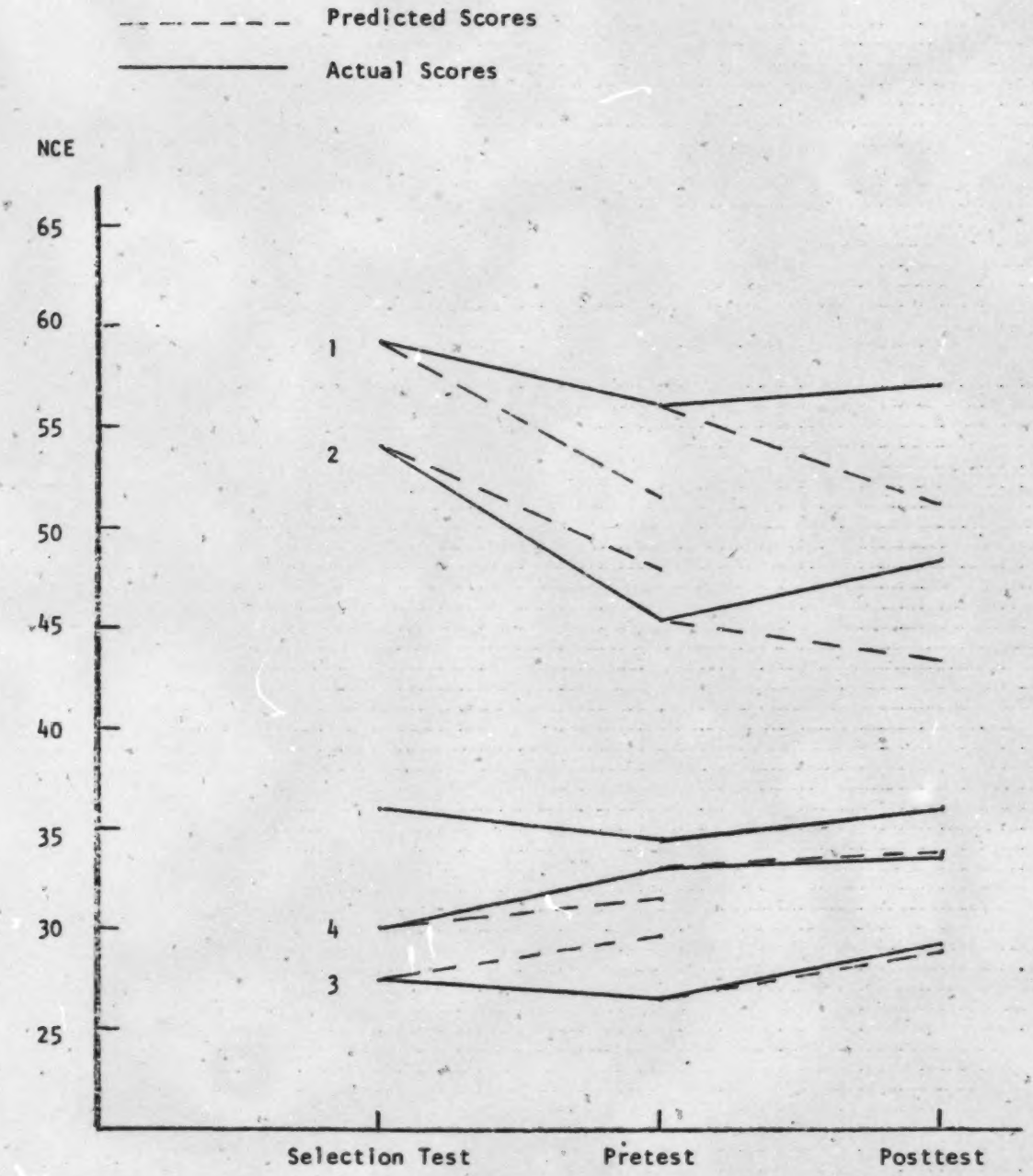


Figure 4
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 6

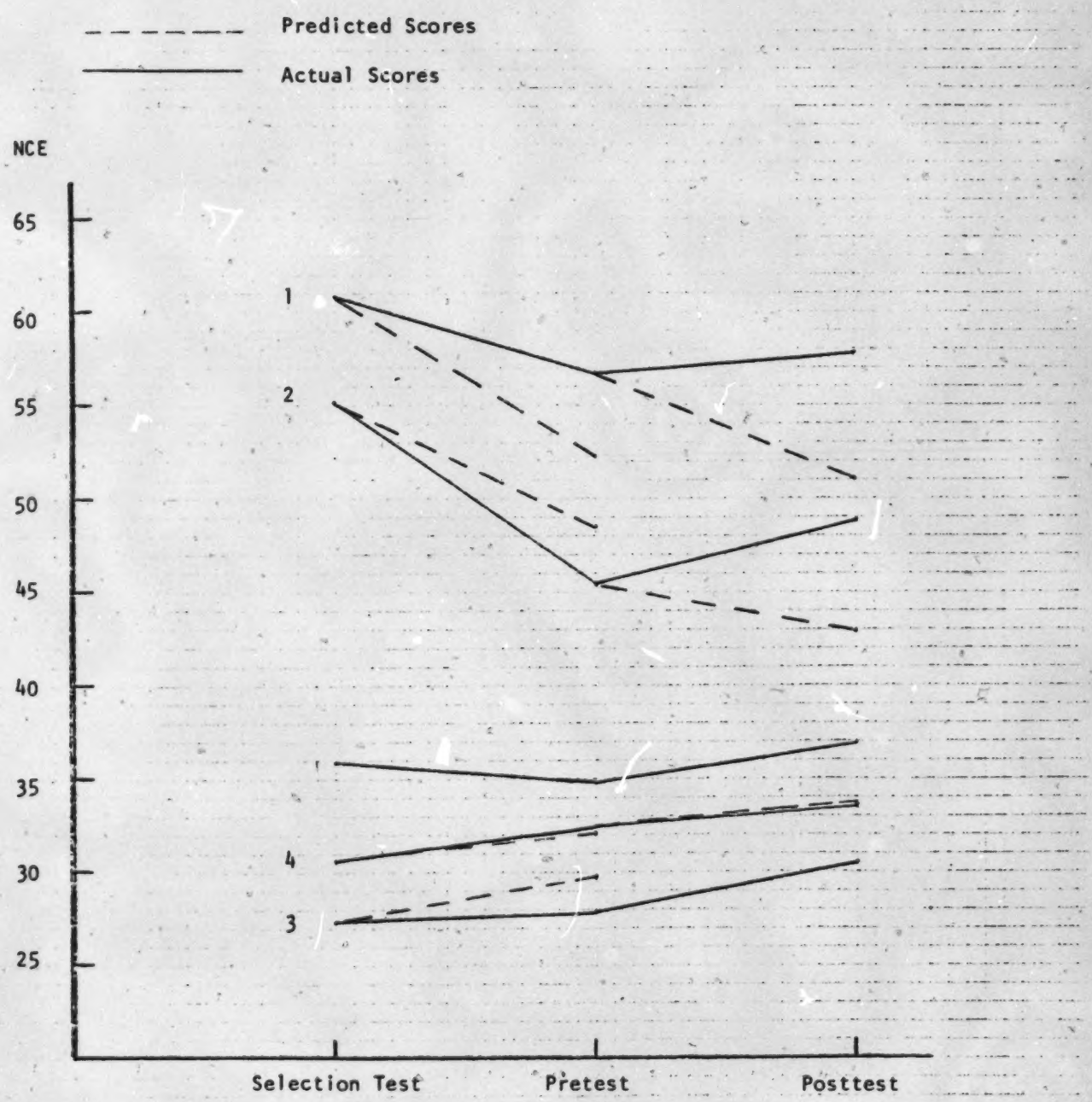


Figure 5
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 7

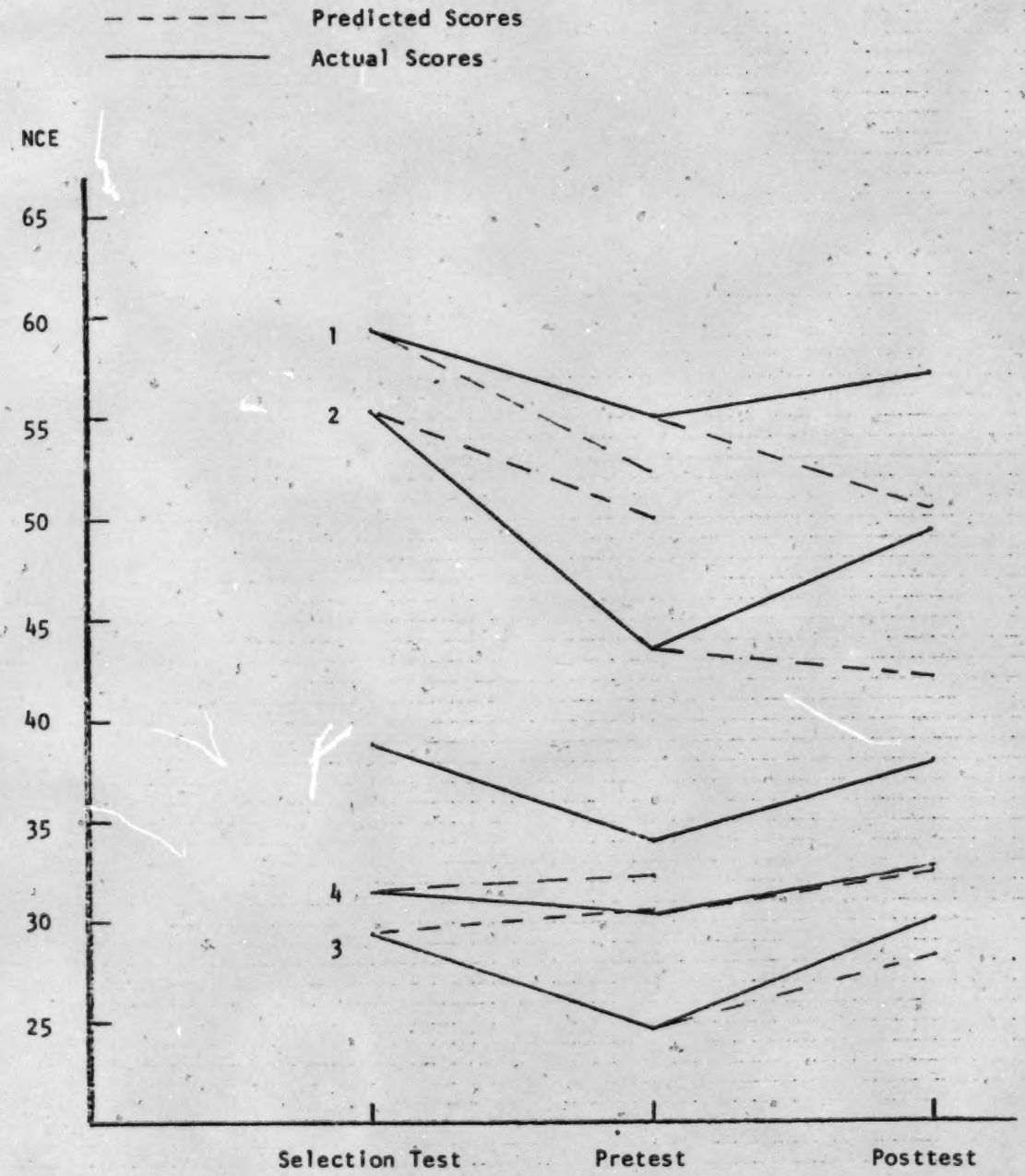


Figure 6
Actual and Predicted
Selection Test, Pretest and Posttest
Reading Means for Samples
of Pupils in Grade 8

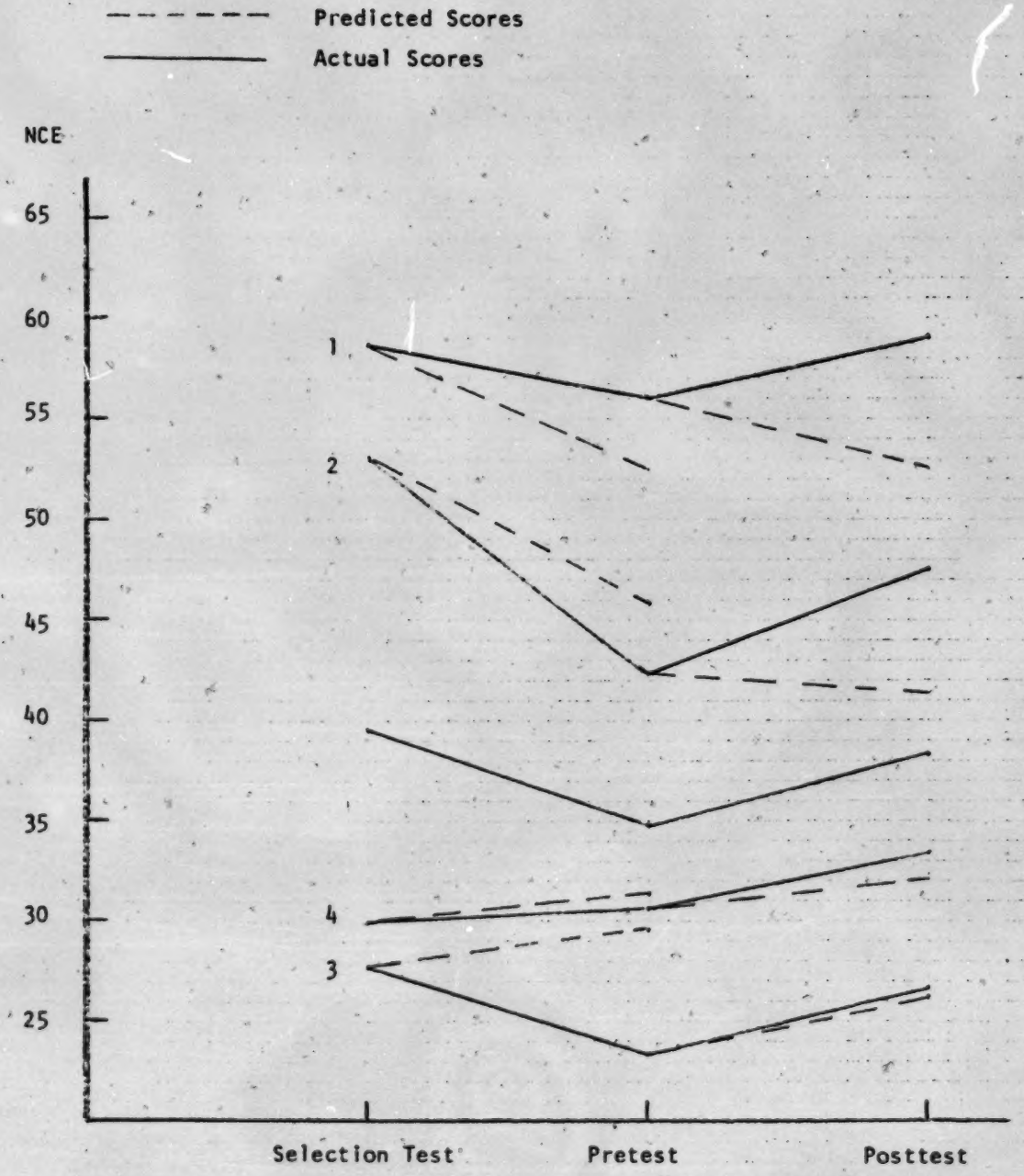


Figure 7
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
of Pupils in Grade 3

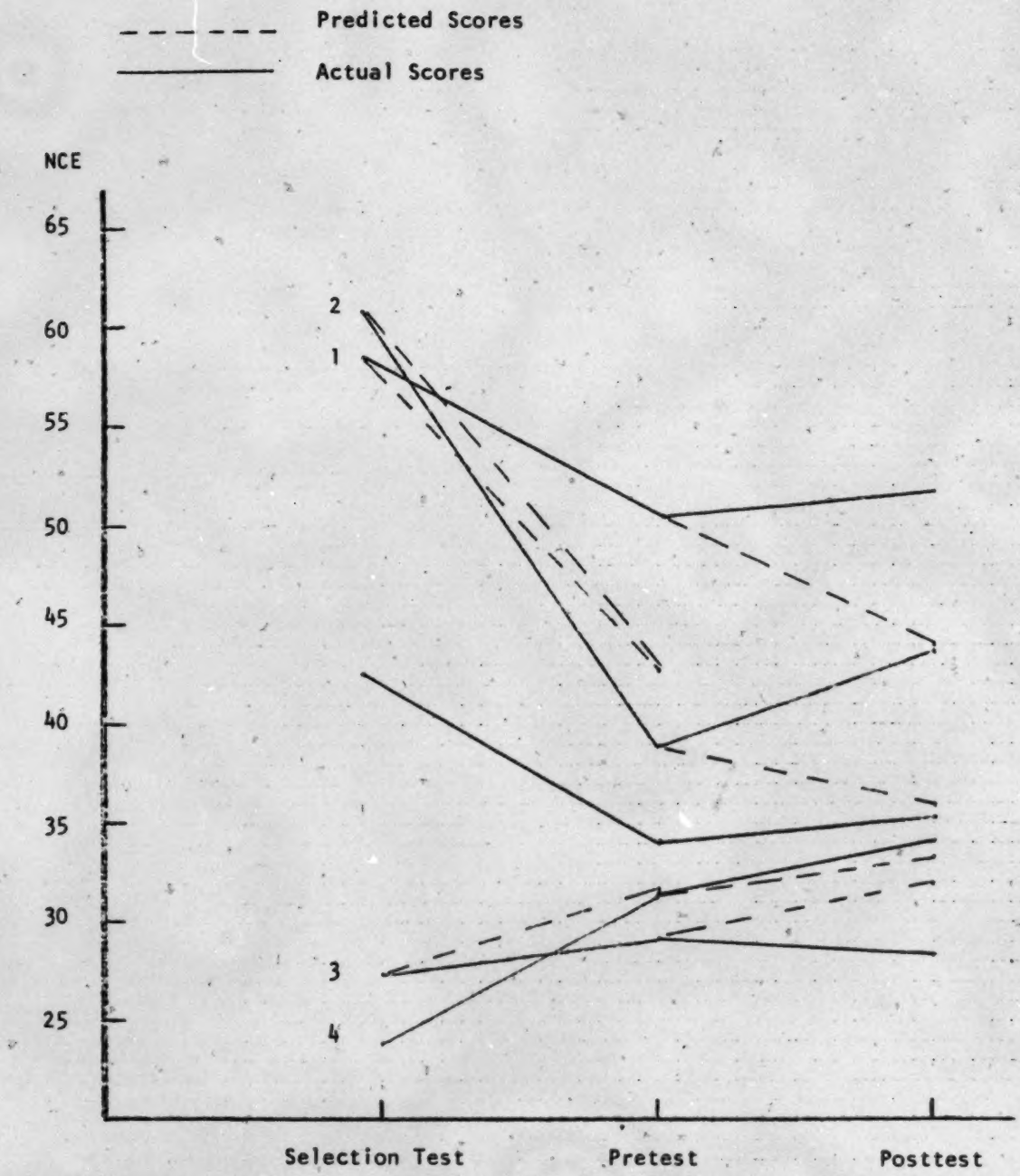


Figure 8
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
Of Pupils in Grade 4

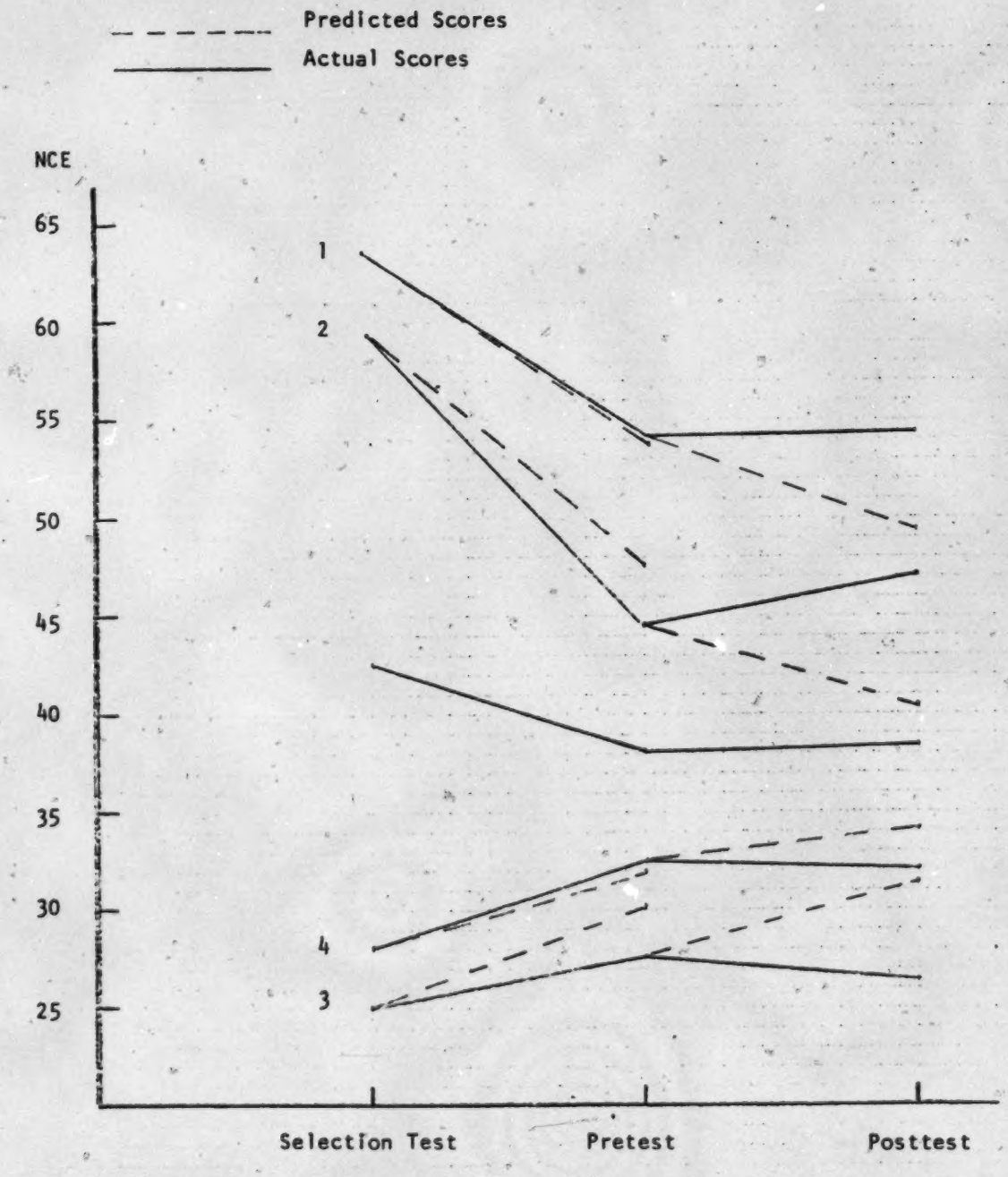


Figure 9
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
of Pupils in Grade 5

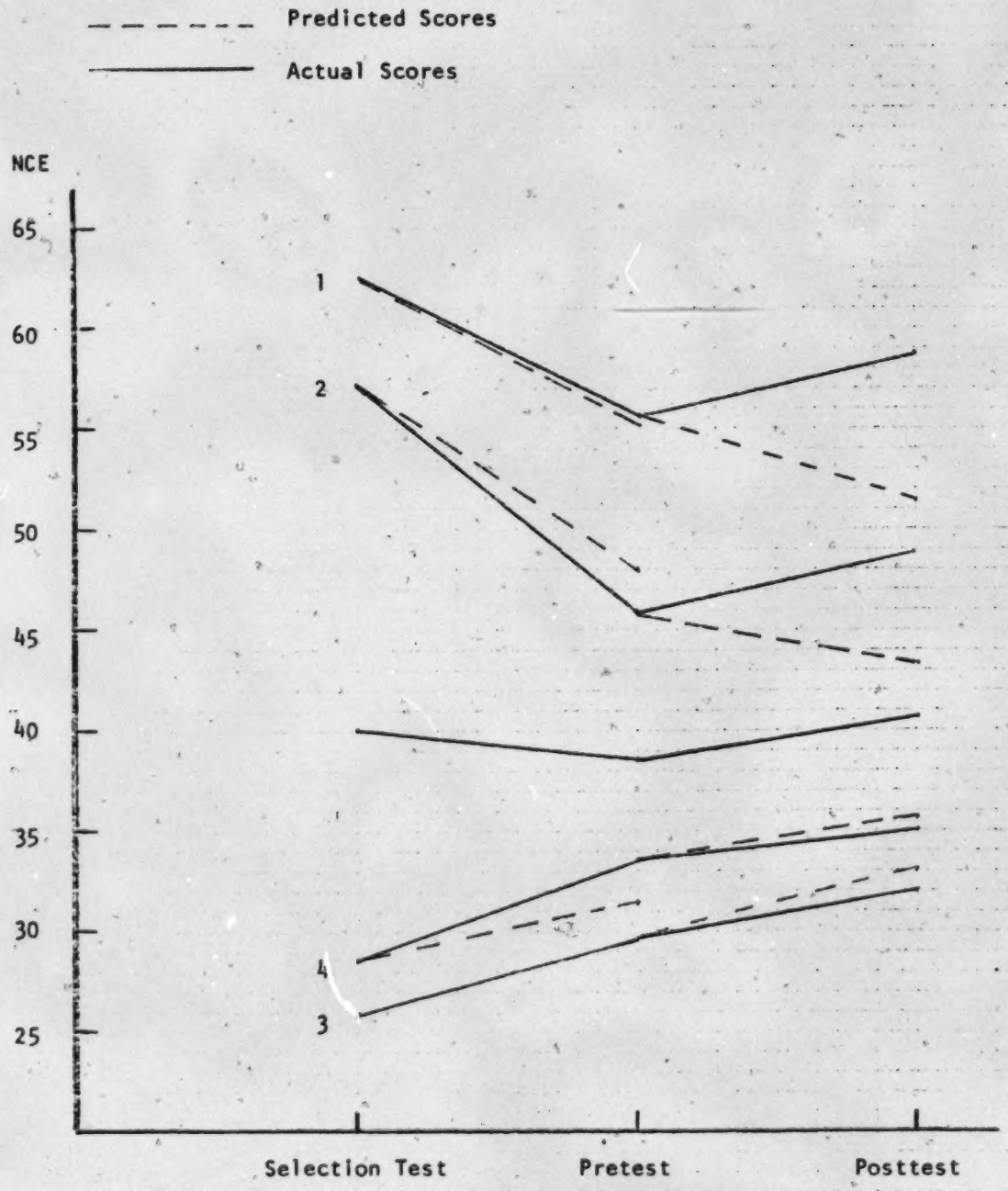


Figure 10
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
of Pupils in Grade 6

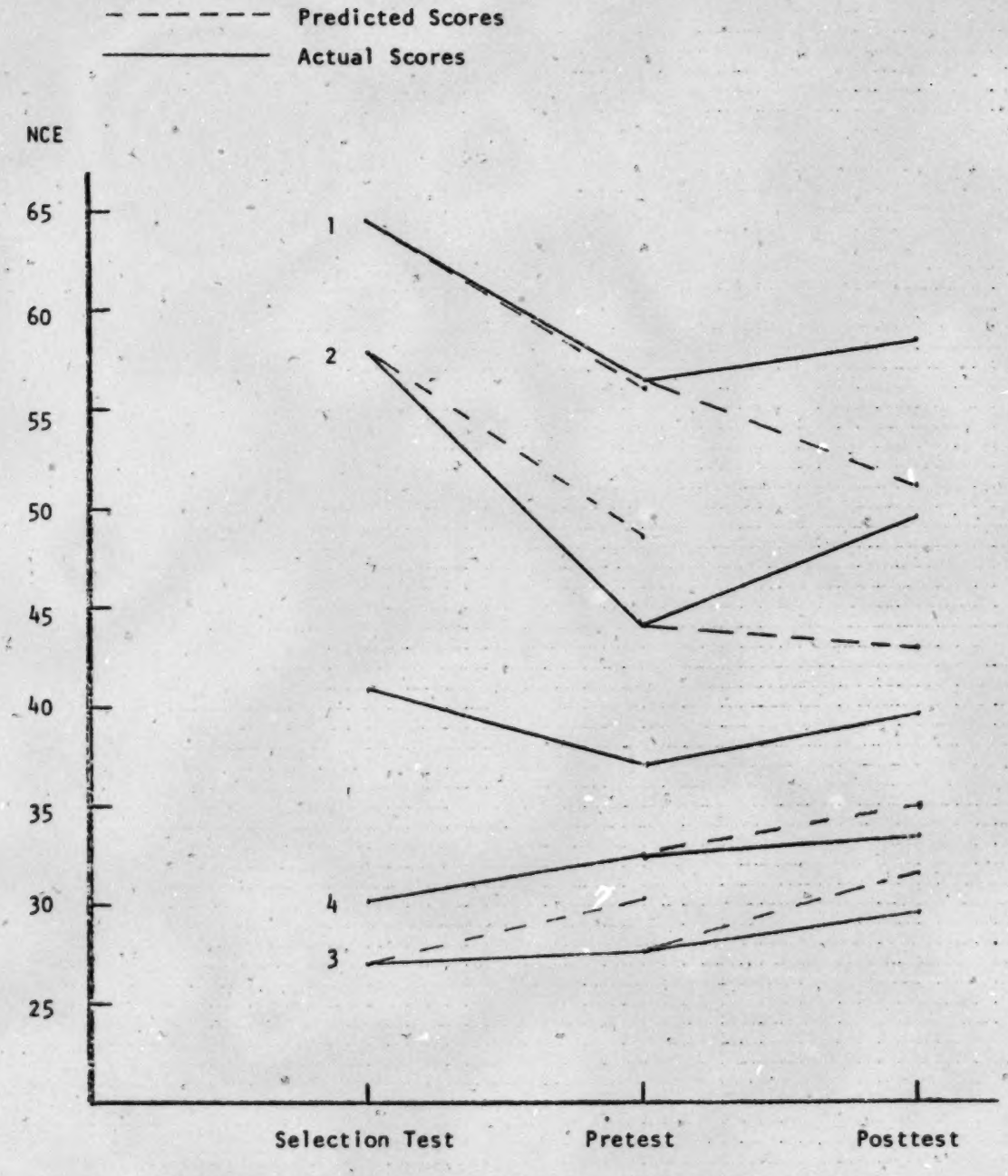


Figure 11
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
of Pupils in Grade 7

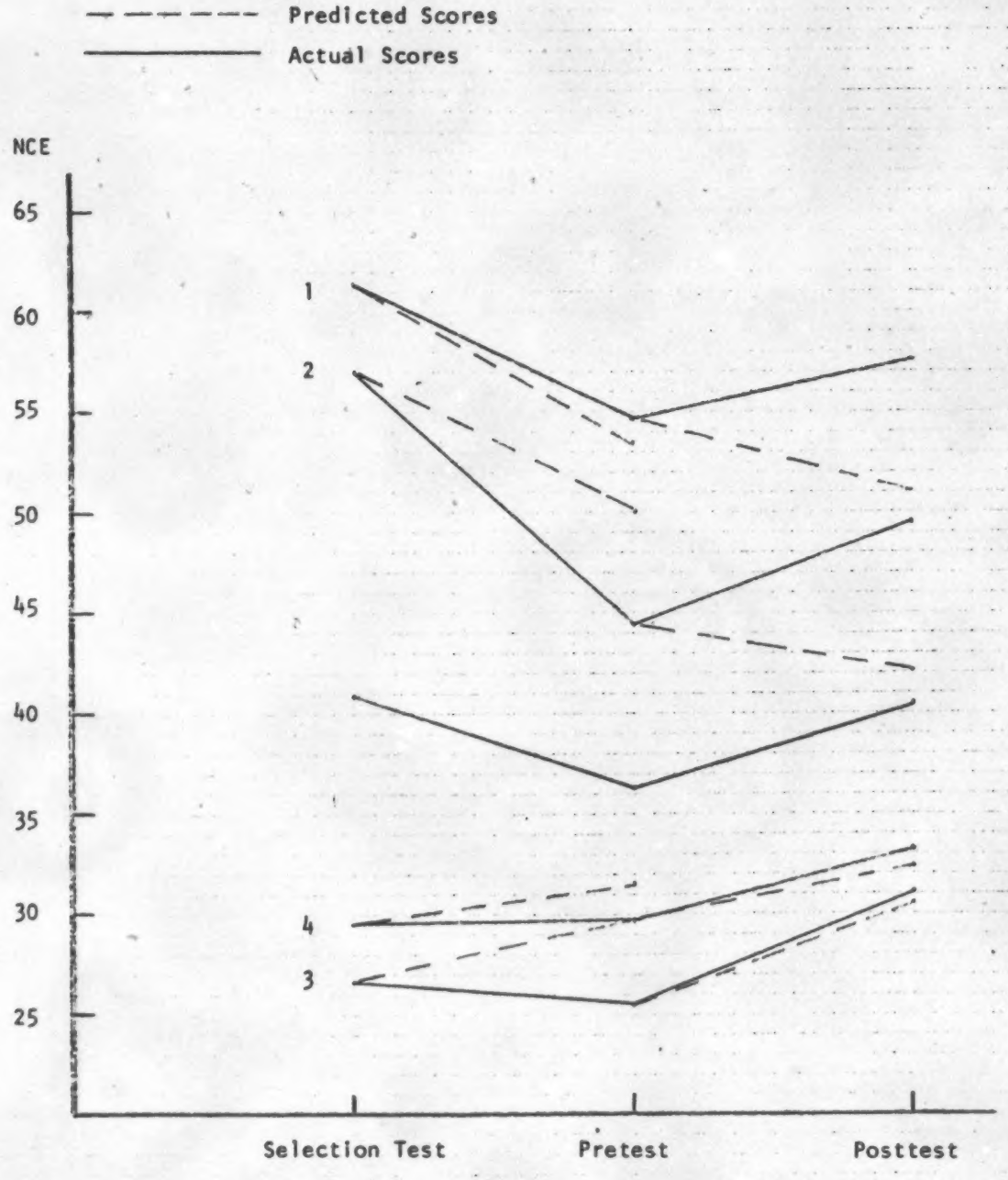
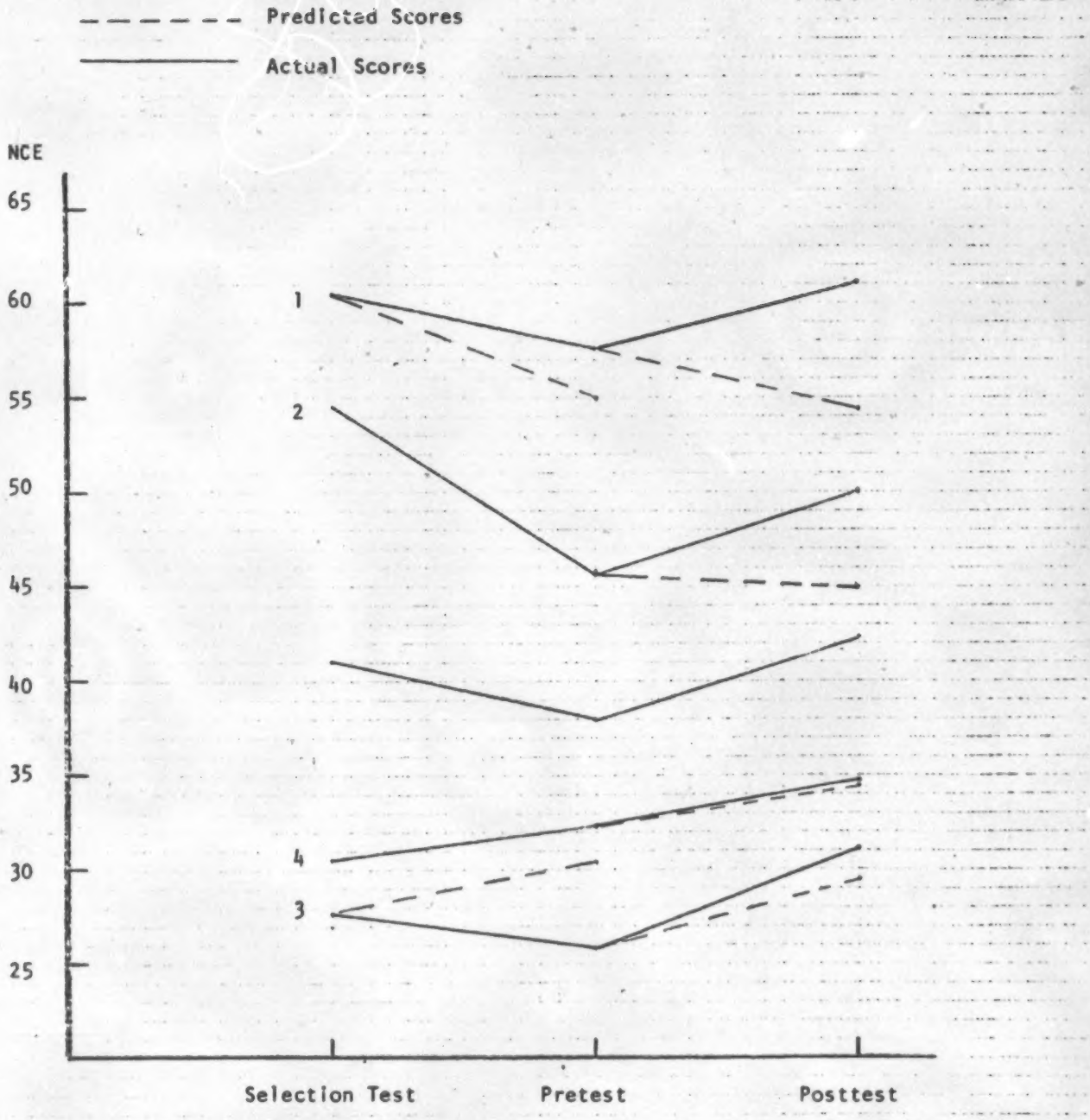


Figure 12
Actual and Predicted
Selection Test, Pretest and Posttest
Arithmetic Means for Samples
of Pupils in Grade 8



Discussion

The data described under Results provide considerable circumstantial evidence that coaching, especially among Title I eligible pupils, occurred during both spring test administrations. All four criteria listed within the hypothesis of this study were confirmed by the results:

1. Fall test performance for the entire tested population was lower at each grade level than was either the prior spring performance or the subsequent test performance. Evidence for this finding can be observed in the selection test, pretest, and posttest population area curves shown in figures 1 through 12.

2. Fall test results for eligible and not identified pupils dropped more than did that of any other group, and more than the population as a whole. Curves for group two show that this criterion was met at every grade level and on both subtests. Further evidence of a dramatic drop in pretest performance as compared to selection test performance for group two is found in tables 1 and 2, which show that virtually half of these pupils became eligible after pretesting, whereas none were identified on the selection test. Proportions of group two which became identified exceeded that of their group one counterparts by approximately 50 percent.

3. Fall test performance for eligible and identified pupils did not change substantially from previous spring levels, nor did it change substantially from that of their Title I ineligible, relatively low scoring counterparts. This can be seen in the proximate observed and predicted curves for groups three and four on figures 1 through 12. It can also be confirmed by the high proportions of these groups who remained identified through pretest and posttest administrations, as reported in tables 1 and 2.

4. Performance on tests subsequent to the previous spring selection test varied significantly from statistical expectations for Title I eligible and not identified pupils. The mean test performance for all other groups did not differ substantially from statistical predictions.

The fact that the above conclusions hold true across six grade levels and two ITBS subtests is strong circumstantial evidence that coaching could have taken place. Of course, the way to become more comfortable with the

position that spring test coaching is a pattern within certain schools or districts would involve some clever research designing and implementing.

However, since decisions which greatly affect the educational programs of literally thousands of pupils are contingent upon the reliability of test results, such a study may well be justified.

At the very least, this study points to a strong need for the school system in which the study took place to revise its Title I pupil selection procedures in such a way as to increase the likelihood that served pupils are really the ones who need Title I services. The results of this study also speak to anyone who attempts to interpret test scores for the school system, especially those whose interest it is to evaluate changes in scores between test administrations.

References

Campbell, Donald T., Reforms as experiments. Reprinted with revisions from American Psychologist, 24:4, Strunening and Guttentag (Eds.), Handbook of Evaluation Research, Vol. I., Sage Publications, 1975.

Campbell, D. T, and A. Erlebacher, How regression artifacts in quasi-experimental evaluation can mistakenly make compensatory education look harmful. Reprinted from Helmmuth, J. (Ed.), Compensatory Education: A National Debate, Vol. III of The Disadvantaged Child, New York: Brunner/Hazel, 1970. Struening and Guttentag (Eds.), Handbook of Evaluation Research, Vol. I., Sage Publications, 1975.

Campbell, D. T. and J. C. Stanley, Experimental and Quasi-Experimental Designs for Research. Reprinted from Handbook of Research on Teaching, the American Educational Research Association, 1963.

Hopkins, Kenneth D., Regression and the matching fallacy in quasi-experimental research. Unpublished manuscript. Laboratory of Educational Research, University of Colorado, circa 1977.

Tallmadge, G. Kasten, The regression effect. E.S.E.A. Title I Evaluation and Reporting System, Technical Paper No. 3., RMC Research Corporation, October, 1976.